

11.3 Describing and Analyzing Data





Goals for the Day

1

Measures of
Center

2

Measures of
Spread

3

Using your
Calculator

4

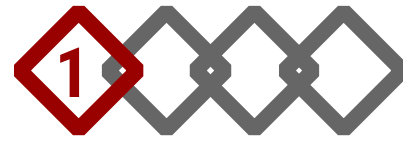
Measures of
Relative
Position

1

Measures of Center



Mean (Average Value)



Simple, arithmetic average of the data.

▷ Sum all numbers and divide by the sample size (n).

Same calculation for the population and sample mean (just different notation).

▷ Sample mean = \bar{x} (pronounced "x-bar")

▷ Population mean = μ (Greek letter mu)

Mean is NOT a resistant measure.

▷ This means it is heavily affected by outliers.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Example 1 – Mean

Data: 1, 5, 2, 9, 3

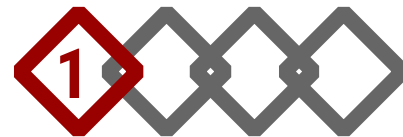
$$\bar{x} = \frac{1 + 5 + 2 + 9 + 3}{5} = 4$$

Now change 3 to 30

New $\bar{x} = 9.4$



Median (Middle Value)



■ The middle value in an ordered list.

■ Median IS a resistant measure.

▷ NOT affected by outliers.

Example 2 – Median

Case 1 – Odd n

7 Obs: 10, 5, 6, 1, 3, 9, 8

Sorted: ~~1~~, ~~3~~, ~~5~~, 6, ~~8~~, ~~9~~, ~~10~~

$$Med = 6$$

Case 2 – Even n

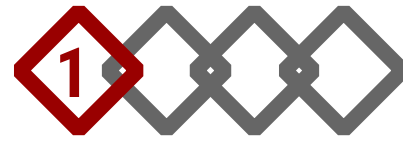
8 Obs: 10, 5, 6, 1, 3, 9, 8, 3

Sorted: ~~1~~, ~~3~~, ~~3~~, 5, 6, ~~8~~, ~~9~~, ~~10~~

$$Med = \frac{5 + 6}{2} = 5.5$$



Mode (Most Common Value)



■ The most frequently occurring value(s).

- ▷ Unimodal data has one mode.
- ▷ Bimodal data has 2 modes.
- ▷ Multimodal data has more than 2 modes.
- ▷ Can be no modes (every value is distinct).

■ This is the only measure of center that can be used with categorical data.

- ▷ Ex) Most common favorite color (can't average this)

Example 2 – Median

Case 2

8 Obs: 10, 5, 6, 1, 3, 9, 8, 3

Mode = 3 (twice)

2

Measures of Spread



Range



- Range = Max - Min
- Gives idea of the entire "range" of values, how much distance do they span in total.
- Ex) Case 2: Range = $10 - 1 = 9$

Example 2 – Median

Case 2

8 Obs: 10, 5, 6, 1, 3, 9, 8, 3

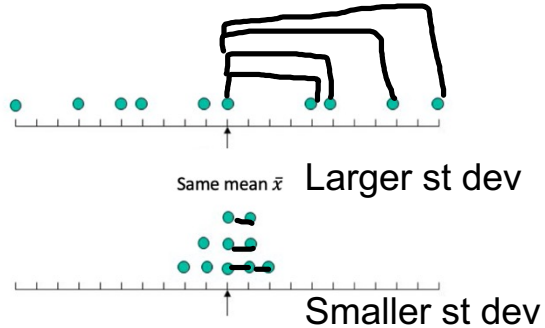




Standard Deviation



Complex formula that measures the average distance that each data point is from the mean.



$$\text{Sample Standard Deviation} = S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

$$\text{Population Standard Deviation} = \sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

(Greek letter sigma)

3

Using your Calculator

Using Your Calculator



Steps for the TI-83/84

1. Enter data: STAT → Edit → Enter data in L_1
(Demo dataset: 10, 23, 4, 6, 9, 3, 15, 6)
2. Calculate: STAT → CALC → 1-Var Stats
 - a) List is L_1 .
 - b) Leave FreqList blank.
 - c) Calculate!

Steps for the TI-30XS MultiView

1. Data → Enter data in L_1
2. 2nd → stat → 1-Var Stats
 - a) DATA: L_1
 - b) FRQ: ONE
 - c) CALC

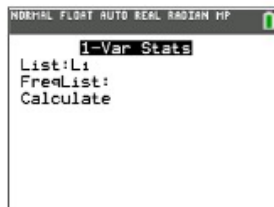
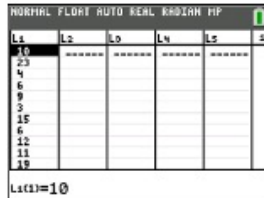
Steps for the TI-30 XIIS

1. 2nd → STAT → 1-VAR (Enter)
2. DATA
 - $X_1 = \#$ (scroll down)
 - FRQ = 1 (for ALL X_s , scroll down)
 - $X_2 = \#$ (scroll down)
 - ...
3. STATVAR (scroll across)
4. To exit this menu: 2nd → EXIT STAT → Y

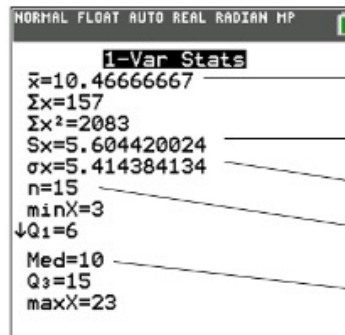
Inputs

Data here:

10, 23, 4, 6, 9,
3, 15, 6, 12, 11,
19, 10, 6, 8, 15



Results



Does not give median

Mean

Sample st dev

Population st dev

Sample size

Median



Using Your Calculator Example

Example 3: Find the mean, median, mode and sample standard deviation of the following dataset.

▷ Data (7 obs): 35, 70, 31, 37, 65, 38, 38

Results

Mean $\bar{x} = 44.86$

Sample st dev: $S_x = 15.72$

Pop st dev $\sigma_x = 14.55$

Med = 38

Other Considerations



Outliers

- ▶ Data values that are extreme when compared to the rest of the data.
- ▶ Can significantly impact measures of center and spread.

Example 3:

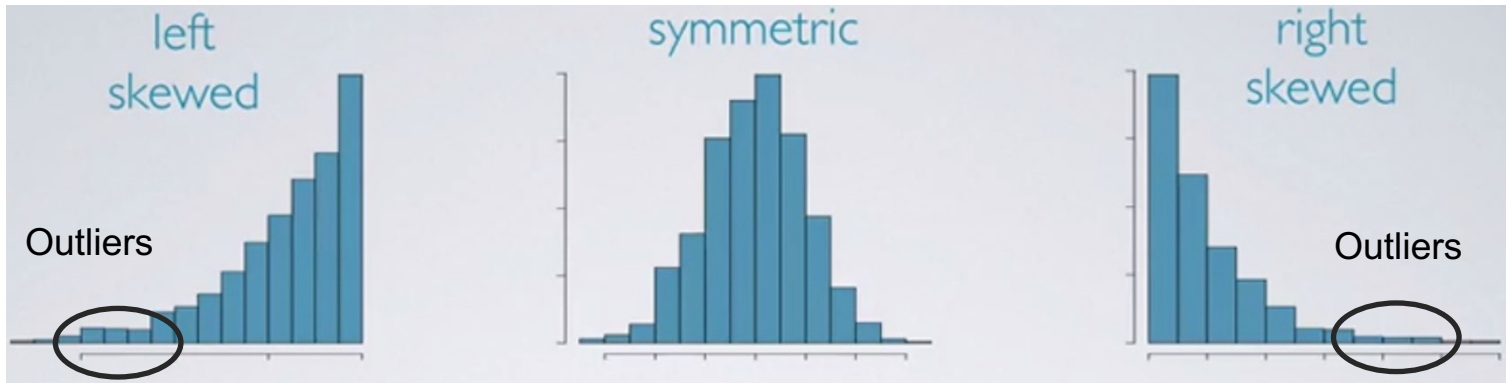
- Data (7 obs): 31, 35, 37, 38, 38, 65, 70

Outliers

Other Considerations



Types of Distributions



Best measure of center: Median

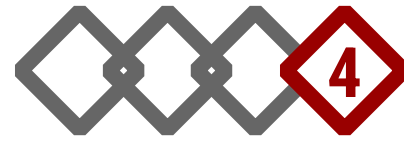
Mean

Median

4

Measures of Relative Position

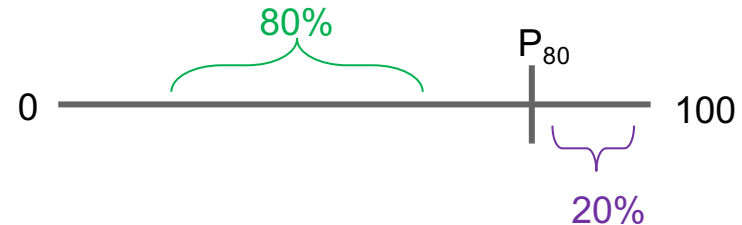
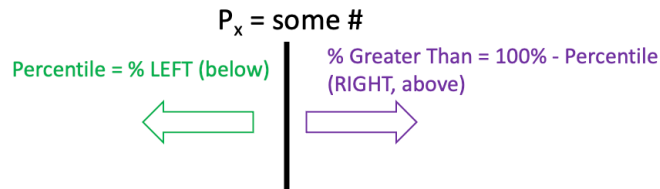
Percentiles



■ A **percentile** tells you the percent of observations/individuals you are higher than.

- Interpreting example: You are told you scored in the 90th percentile on GRE. This means you have a score that is higher than 90% of all others that took the test.
- Range from 0th to 100th percentile!
- There is complement aspect to percentiles as well; for example, if you are the 80th percentile, there is 20% greater than you!

■ Best way to remember!



- Notation: Xth Percentile = P_x

5-Number Summary and Boxplots

Quartiles are specific percentiles.

- ▷ Q_1 is the 25th Percentile.
- ▷ Q_3 is the 75th Percentile.
- ▷ Q_2 is the 50th Percentile = Median.

Inner Quartile Range (IQR)

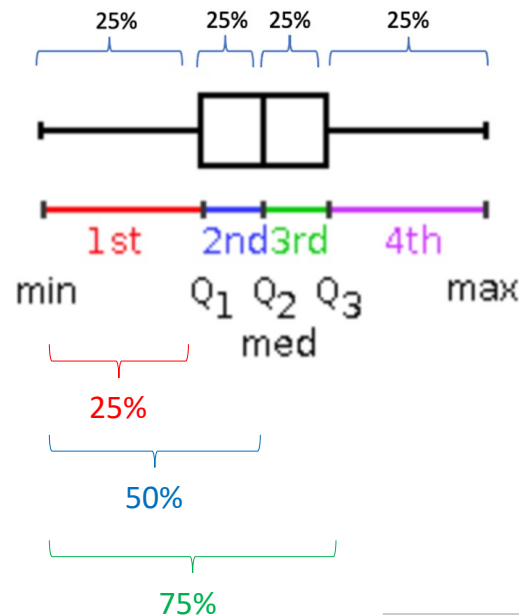
- ▷ Another measure of variation, less informative than the standard deviation.
- ▷ Uses quartiles to measure how far data is spread out around the median. Specifically, it measures the range of the middle 50% of the data

▷ $IQR = Q_3 - Q_1$

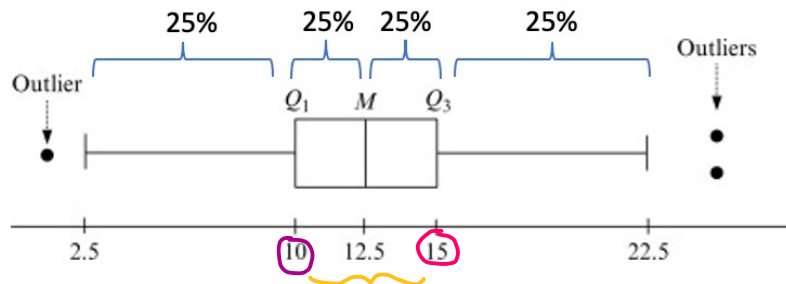
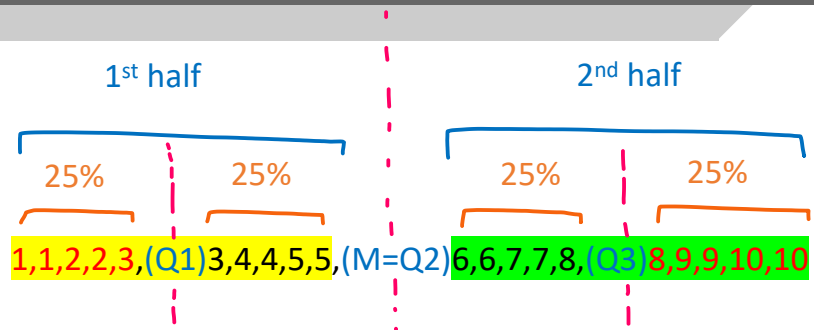
- ▷ Visualized very well in boxplots! It is the length of the box!

5-number summary

- ▷ $\text{Min, } Q_1, \text{ Med, } Q_3, \text{ Max}$ → Points of a boxplot



5-Number Summary and Boxplots

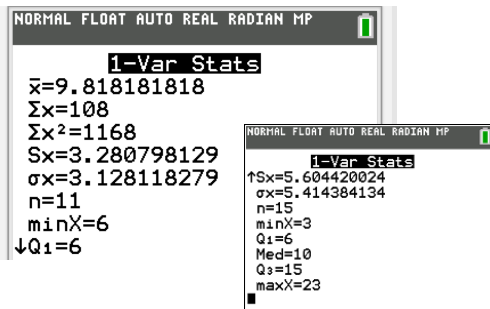


$$IQR = Q_3 - Q_1 = 15 - 10 = 5$$

Examples

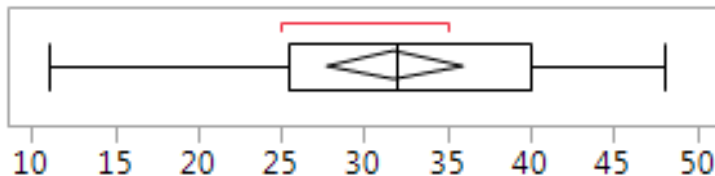
Example 4:

- a) Using this output from a 1-Var Stat, what is the IQR?



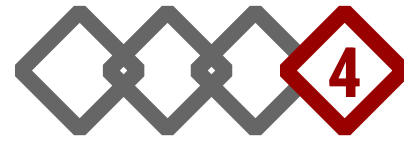
$$IQR = 9$$

- b) Find the IQR from this boxplot.



$$IQR = 15$$

Examples



Example 5:

a) Calculate the 5-number summary of the following dataset (20 numbers):

38, 33, 5, 5, 47, 29, 24, 42, 3, 18, 30, 46, 25, 44, 40, 42, 39, 44, 29, 13

5-number summary

$Min = 3$

$Q_1 = 15.5$

$Med = 29.5$

$Q_3 = 41$

$Max = 47$

b) Draw a boxplot based on the 5-number summary from (a).

